Logic and probabilities for recommender systems

Tanel Tammet
Ago Luberg, Kalle Tomingas, Priit Järv
Tallinn Uni of Tech

Contents

- Tourism recommenders as a case study
- Input and output
- Which probabilities we need?
- Simple layered semantics
- Cumulating evidence
- Rankings via meta-logical calculations

Recommender systems?

- Several historical „expert systems“ were recommender systems (medicine etc)
- Google is a popularity-focused recommender
- Social network systems are recommender systems: recommend news items and possible friends and topics
- The wealth of data available online

Two main recommender types

- Collaborative filtering
- Rule-based, also called content-based

Our tourism recommender project

- http://www.sightsplanner.com
- http://www.sightsmap.com

Firefox

Sightsplanner

http://tallinn.sightsplanner.com/

android 2.2 start without butt

VISITOR | PROFESSIONALS | MEETING PLANNER | TALLINN.EE | TALLINN.MOBI

EESTI | ENGLISH 27.03.2011 | 12:57

Walk in the city

Nice walk around the medieval Old town, history and cafes

Choose

Active life

Great events, spiced with sports and outdoor activities

Choose

A night in the town

Clubs, pubs, party and dancing all night long!

Choose

Culture

Beautiful architecture, amazing paintings and concerts

Choose

Get your plan

or see what's popular

	Less	More
+ Events		
+ Museums & Arts		
+ Architecture & City		
+ Eating out		
+ Shopping		
+ Sports & Outdoor		
+ Bars & Nightlife		

Time 14:00 28.03.2011 Duration 1h 2h 3h 6h 8h by car on foot

Firefox

Sightsplanner

http://tallinn.sightsplanner.com/

android 2.2 start without butt

VISITOR | PROFESSIONALS | MEETING PLANNER | TALLINN.EE | TALLINN.MOBI

EESTI | ENGLISH 27.03.2011 | 12:57

Walk in the city

Nice walk around the medieval Old town, history and cafes

Choose

Active life

Great events, spiced with sports and outdoor activities

Choose

A night in the town

Clubs, pubs, party and dancing all night long!

Choose

Culture

Beautiful architecture, amazing paintings and concerts

Choose

Get your plan

or see what's popular

Less More

+ Events

- Museums & Arts

Modern art gallery
Installation, Happening
Old history museum
Science museum

Classical art exhibition
Other art
Contemporary history
Nature museum

+ Architecture & City

- Eating out

Traditional restaurant
European restaurant
Fine dining

Asian restaurant
Other restaurant
Pub

Time 14:00 29.03.2011 Duration 1h 2h 3h 6h 8h by car on foot

1. Estonian History Museum - Great Guild Hall Read more Locate on map

Museums and Arts | Architecture and City | Old history museum | Medieval architecture

In the course of time, the Great Guild Hall has played an important role in the life of the city. The permanent exhibition of the History Museum located in the building introduces Estonia's history...

Arrive 14:05
Stay
45
minutes
Remove from selection

Walk 5 minutes

2. Katariina käik (St Catherine's Passage) Read more Locate on map

Landmark | Architecture and City | Medieval architecture

Vene and Müürivahe Streets are connected by Katariina käik. You can see the remaining parts of St Catherine's Church in its northern end The southern part of the passage is lined by residenti...

Arrive 14:55
Stay
20
minutes
Remove from selection

Walk 5 minutes

3. Olde Hansa Read more Locate on map

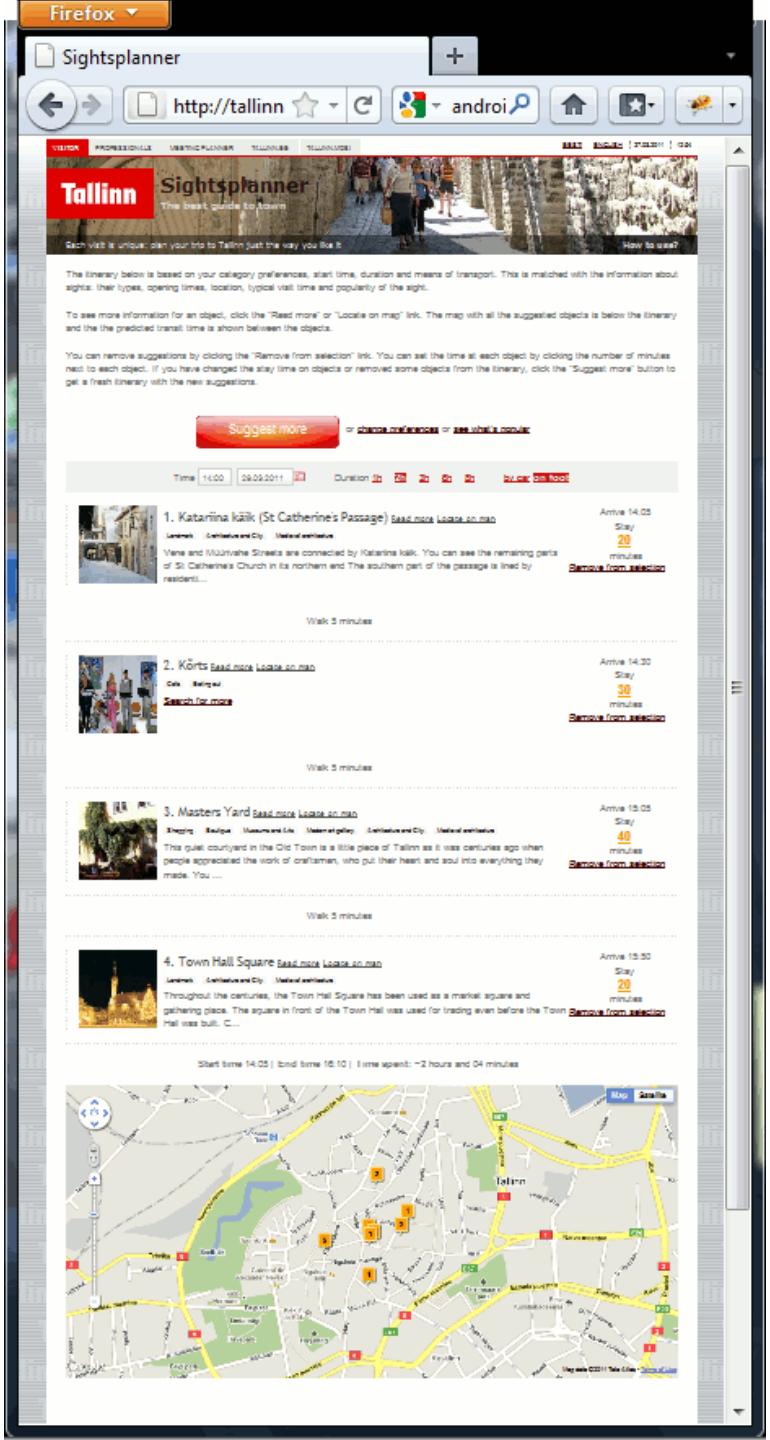
Eating out | Traditional restaurant

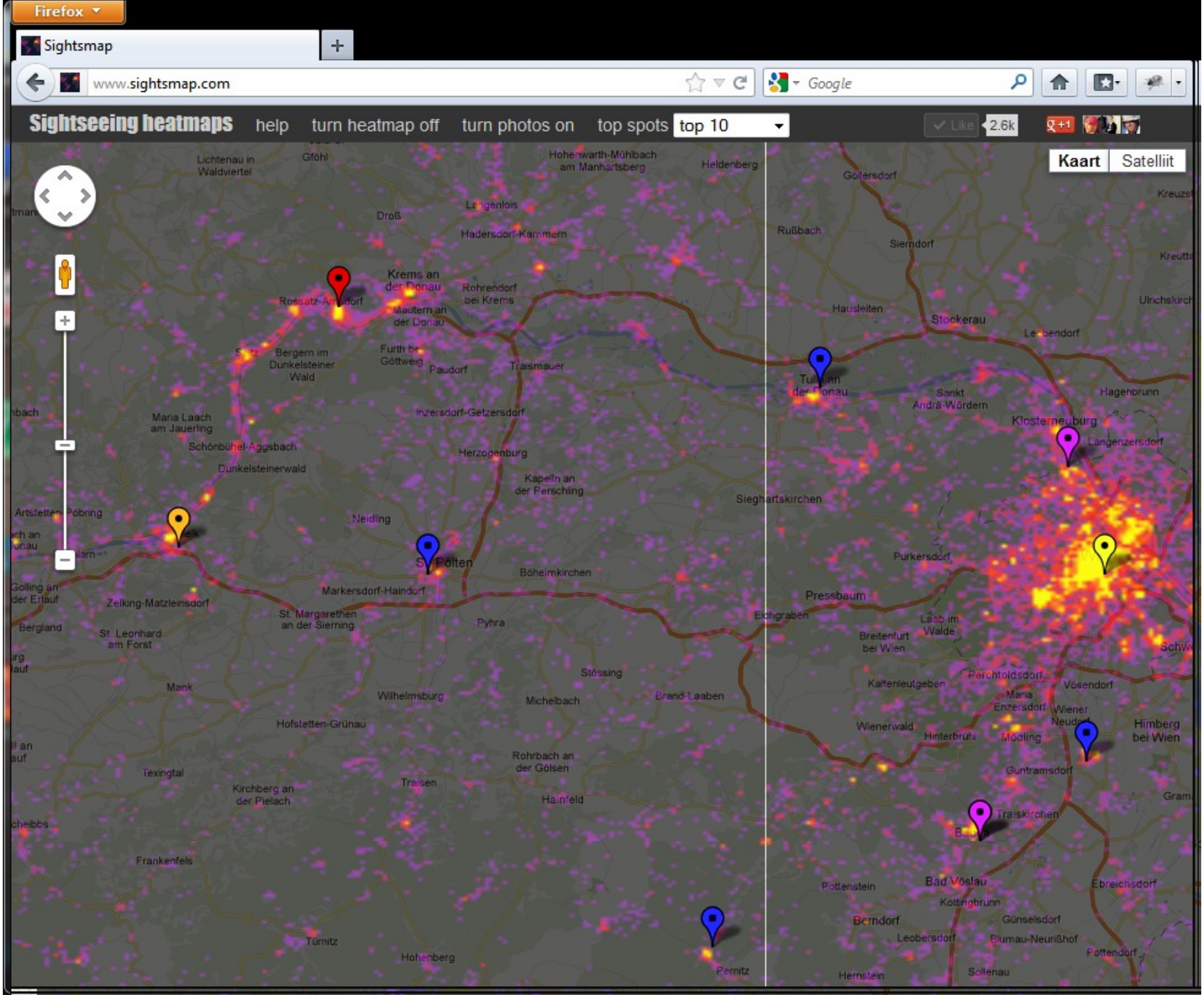
Olde Hansa Restaurant

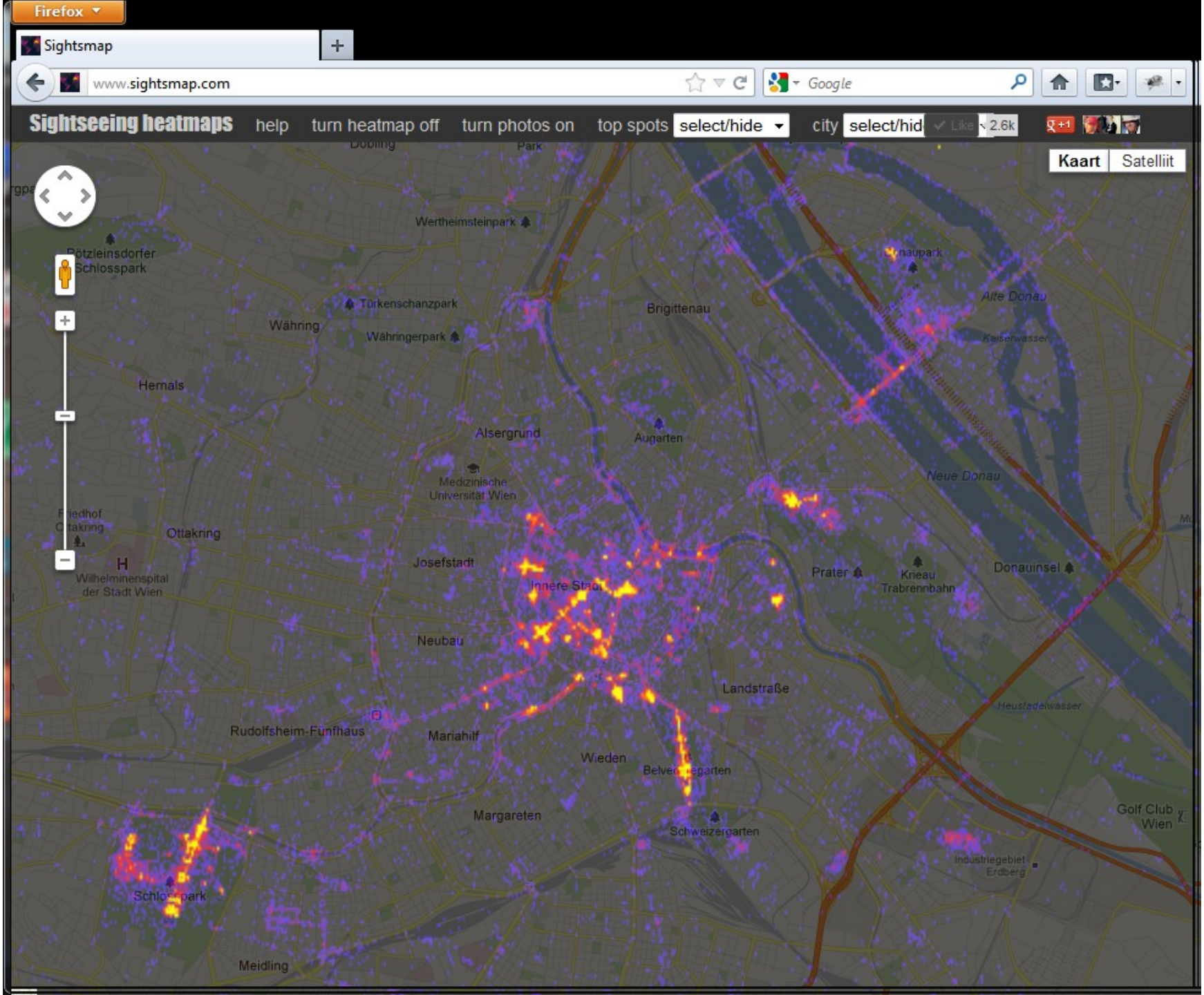
See more

Arrive 15:20
Stay
60
minutes
Remove from selection

Walk 10 minutes


Firefox
Sightsplanner
http://tallinn
Tallinn
Sightsplanner
The best guide to town
Each visit is unique: plan your trip to Tallinn just the way you like it
How to use?
The itinerary below is based on your category preferences, start time, duration and means of transport. This is matched with the information about sights: their types, opening times, location, typical visit time and popularity of the sight.
To see more information for an object, click the "Read more" or "Locate on map" link. The map with all the suggested objects is below the itinerary and the the predicted transit time is shown between the objects.
You can remove suggestions by clicking the "Remove from selection" link. You can set the time at each object by clicking the number of minutes next to each object. If you have changed the stay time on objects or removed some objects from the itinerary, click the "Suggest more" button to get a fresh itinerary with the new suggestions.
Suggest more
or change preferences or see what's popular
Time 14:00 28.09.2011 Duration 1h 2h 3h 4h 5h by car on foot
1. Katariina käik (St Catherine's Passage) Read more Locate on map
Vene and Müürivahe Streets are connected by Katariina käik. You can see the remaining parts of St Catherine's Church in its northern and The southern part of the passage is lined by residenti...
Arrive 14:05
Stay
20
minutes
Remove from selection
Walk 5 minutes
2. Kõrts Read more Locate on map
Search for more
Arrive 14:30
Stay
30
minutes
Remove from selection
Walk 5 minutes
3. Masters Yard Read more Locate on map
This quiet courtyard in the Old Town is a little piece of Tallinn as it was centuries ago when people appreciated the work of craftsmen, who put their heart and soul into everything they made. You ...
Arrive 15:05
Stay
40
minutes
Remove from selection
Walk 5 minutes
4. Town Hall Square Read more Locate on map
Throughout the centuries, the Town Hall Square has been used as a market square and gathering place. The square in front of the Town Hal was used for trading even before the Town Hall was built. C...
Arrive 15:50
Stay
20
minutes
Remove from selection
Start time 14:05 | End time 18:10 | Time spent: ~2 hours and 04 minutes
Map
Satellite
Tallinn


Firefox
Sightsmap
www.sightsmap.com
Google
Sightseeing heatmaps
help
turn heatmap off
turn photos on
top spots
top 10
Like
2.6k
Kaart
Satelliit
Lichtenau in Waldviertel
Gföhl
Hohenwarth-Mühlbach am Manhartsberg
Heldenberg
Göllersdorf
Droß
Langenlois
Hadersdorf-Kammern
Rußbach
Sierndorf
Krems an der Donau
Rohrendorf bei Krems
Rossatz-Arnsdorf
Mautern an der Donau
Hausleiten
Stockerau
Ulrichskirchen
Furth bei Göttweig
Bergern im Dunkelsteiner Wald
Paudorf
Traismauer
Tulln an der Donau
Sankt Andrä-Wördern
Hagenbrunn
Klosterneuburg
Langenzersdorf
Maria Laach am Jauerling
Inzersdorf-Getzersdorf
Schönbühel-Aggsbach
Herzogenburg
Dunkelsteinerwald
Kapelln an der Perschling
Sieghartskirchen
Neidling
St. Pölten
Purkersdorf
Böheimkirchen
Markersdorf-Haindorf
Pressbaum
Zelking-Matzleinsdorf
St. Margarethen an der Sierning
Eichgraben
Pyhra
Breitenfurt bei Wien
St. Leonhard am Forst
Stössing
Mank
Wilhelmsburg
Michelbach
Brand-Laaben
Kaltenleutgeben
Perchtoldsdorf
Vösendorf
Maria Enzersdorf
Wiener Neudorf
Wienerwald
Hinterbrühl
Mödling
Himberg bei Wien
Hofstetten-Grünau
Rohrbach an der Gölsen
Texingtal
Traisen
Kirchberg an der Pielach
Hainfeld
Güntramsdorf
Traiskirchen
Frankenfels
Pottenstein
Bad Vöslau
Ebreichsdorf
Kottingbrunn
Berndorf
Günselsdorf
Leobersdorf
Baumau-Neurißhof
Pottendorf
Türnitz
Hohenberg
Pernitz
Hernstein
Sollenau


Firefox
Sightsmap
www.sightsmap.com
Google
Sightseeing heatmaps
help
turn heatmap off
turn photos on
top spots
select/hide
city
select/hid
Like
2.6k
+1
Kaart
Satelliit
Wertheimsteinpark
Pötzleinsdorfer Schlosspark
Türkenschanzpark
Währing
Währingerpark
Brigittenau
Alte Donau
Kaiserwasser
Hernals
Alsergrund
Augarten
Medizinische Universität Wien
Neue Donau
Ottakring
Josefstadt
Innere Stadt
Prater
Krieau Trabrennbahn
Donauinsel
Wilhelminenspital der Stadt Wien
Neubau
Landstraße
Heustadelwasser
Rudolfsheim-Fünfhaus
Mariahilf
Wieden
Margareten
Schweizergarten
Golf Club Wien
Industriegebiet Erdberg
Meidling

Input 1

- User interests:

likes(john,nightlife,0.6)

likes(john,sports,0.8)

likes(john,music,0.7)

likes(john,heavymetal,0.9)

dislikes(john,classicalmusic,0.9)

Input 2

- Object properties:

type(omalley,bar,0.9)

activity(omalley,footballwatching,0.7)

popularity(omalley,1000)

type(crown,restaurant,1.0)

activity(crown,heavymetal,0.8)

popularity(crown,1500)

opentime(crown,12.00,0.9)

Input 3

- Knowledge about the world:

type(X,church,M) -> type(X,architecture,M*0.9)
type(X,bar,M) -> type(X,drinkingplace,M)
type(X,restaurant,M) -> type(X,drinkingplace,M*0.7)
activity(X,footballwatching,M) -> activity(X,sports,M)

type(X,fastfood,M) -> visitminutes(X,20,0.8*M)
type(X,bar,M) & M>0.75 -> openat12(X,0.85)

description(X,S) &
contains_str(S ,“paintings“) &
contains_str(S ,“gallery“) ->
type(X,artcollection,0.8)

Output

- Recommendations: numerical ranks for all tourism objects:

  rank(john,omalley,0.6)

  rank(john,crown,0.5)

Reasoning tasks

- Object identities: are two objects A and B obtained from different sources actually equal?
- Object types from content: using title, abstract, source etc, calculate wheather the object is a city, a castle, a church, medieval, modern, a drama play, a classical music concert, a rock concert, ...
- Generalised object types: if we know that an object is a bar (with some confidence X), then it is also a nightlife spot (with some confidence Y)
- Additional properties like time of visit,

Probabilities?

There is a large number of probability-oriented theories and several reasoning systems, yet no “mainstream” probabilistic rule-based derivation algorithms exist

Fuzzy logic, probabilistic logic, Bayes networks, ...

Probabilistic datalog, probabilistic prolog, ...

Mycin, Emycin, Cadiag-2, ...

Goal

Formulate a practical, correct and complete way to use probabilities in rules for the (tourism) recommender context, using object logic.

Metalogic:

0.9: type(X,church) -> type(X,architecture)

0.8: type(X,fastfood) -> visitminutes(X,20)

Object logic:

type(X,church,M) -> type(X,architecture,M*0.9)

type(X,fastfood,M) -> visitminutes(X,20,M*0.8)

Which kinds of probabilities?

Non-strict sets a la „blue“, „large“, ...

Fuzzy logic : p(A v B) = max(p(A),p(B))

0.95: type(X,church) -> type(X,architecture)
0.7: type(X,theatre) -> type(X,architecture)

Incomplete knowledge a la „not sure that“ ...

Probabilistic: p(A v B) = p(A)+p(B) – (p(A)*p(B))

0.8: type(X,bar) -> openat12(X)

Object logic:

type(X,church,M) -> type(X,architecture,M*0.9)

type(X,fastfood,M) -> visitminutes(X,20,M*0.9)

Object logic layers of interpretation

Pred(t): Pred(t) holds.
Pred(t,m): Pred(t) holds with a fuzzy measure at least m.
Pred(t,m,c): With confidence (probability) at least c, Pred(t) holds with at least a fuzzy measure m.
Pred(t,m,c,d): The fact "with confidence (probability) at least c, Pred(t) holds with at least a fuzzy measure m„ holds and depends on the set of clauses d.

Examples

bar(malloy,0.9,1): we are certain that malloy is bar with a fuzzy measure at least 0.9

bar(crown,0.9,0.8): we are 0.8 confident that crown is a bar with a fuzzy measure at least 0.9

Rule examples

bar(X,M,C) & M>L -> openat12(X,1,C*0.8):

when we have confidence C in that X is a bar with a measure M at least L, we are C*0.8 confident that it is open at 12 with a measure 1.

optionally
bar(X,M,C) -> openat12(X,1,M*C*0.8):

example of a sure rule:

bar(X,M,C) -> can_eat_at(X,M*0.5,C):

Fuzzy part is easy

Use your own preferred function f and limits for fuzzy derivation

Pred(X,M1) & Pred(X,M2) -> Pred(X, f(M1,M2))

Pred(X,M) & M>L -> Pred(X, f(M))

Standard derivation rules in resolution hold, nothing is added.

We can enhace subsumption, provided f is monotonic:

Pred(X,M1) subsumes Pred(Y,M2) iff Y=Xs and M1>=M2.

Probabilistic part requires tracking

Recall P(t,M,C,D): C is the probability and D is the set of facts on which the atom depends upon.

Always use rules of form

P(...,D1) & ... & P(...,Dn) & A1 & .... & An ->
P(...,union(D1,...,Dn))

where P atoms do contain probabilities and A1 ... An do not contain probabilities

Multiplying probabilities

Generally the rules should have a form

P1(t1,M1,C1,D1) & ... & Pn(tn,Mn,Cn,Dn) -> P(t,M,f(M1,...,M2),g(C1,...,Cn,D1,...,Dn),union(D1,..., Dn))

- In simple cases g(C1,...,Cn,D1,...,Dn) = C1*...*Cn
- However, if intersection(D1,....,Dn) is not empty, Ci-s corresponding to Di-s with multiple occurrences should be used only once

Cumulating evidence

Use evidence cumulating rule schema:

Pred(X,M1,C1,D1) & Pred(X,M2,C2,D2) &
Empty(Intersection(D1,D2))
->
Pred(X,min(M1,M2),(C1+C2)-
(C1*C2),union(D1,D2))

Cumulating evidence

Example: independent facts

a) bar(X,M,C,D) & M>0.75 -> openat12(X,1,C*0.8,D)
b) intitle(X,"allnight",M,C,D) & M>0.75 ->
  openat12(X,1,C*0.9,D)
c) bar(malloy,1,1,{c}).
d) intitle(malloy,"allnight",1,1,{d}).

a,c: e) openat12(malloy,1,0.8,{c})
b,d: f) openat12(malloy,1,0.9,{d})

giving for our case (0.8+0.9=1.7, 0.8*0.9=0.72,
  1.7-0.72=0.98)
openat12(malloy,1,0.98,{c,d})

Cumulating evidence

Example: dependent facts
f) activity(X,heavymetal,1,1,D) ->
activity(X,music,1,1,D).
g) activity(X,Y,M1,C1,D1) & likes(U,Y,M2,C2,D2) ->

fits(U,X,1,M1*M2*C1*C2,union(D1,D2))
a) likes(john,music,1,0.6,{a})
b) likes(john,heavymetal,1,0.8,{b})
c) activity(crown,heavymetal,1,1,{c}).

c,f: h) activity(crown,music,1,1,{e}).
g,a,h(cf): i) fits(john,crown,1,0.6,{a,c})
g,b,c: j) fits(john,crown,1,0.8,{b,c})

Ranking calculation in meta-logic

- Derive all open-at-time facts.
- Derive all independent addrank facts, using:

  Popularity(X,P) -> addrank(X,pf(P))

  Likes(X,Y,M1) & assoc(Z,Y,M2,C,D) ->
  addrank(X,Z,f(M1,M2,C),D)
  Dislikes(X,Y,M1) & assoc(Z,Y,M2,C,D) ->
  addrank(X,Z,nf(M1,M2,C),D)

- Sum all maximal pos/neg addrank numbers for objects.
- Filter out objects which are open at time.
- Order by rank.

Summary 1

Represent facts as P(t,M,C,D) where:

M- fuzzy measure of P(t) holding

C – confidence as probability of at least P(t,M) holding

D – set of facts on which P(t,M,C) depends

Represent rules as

P1(t1,M1,C1,D1) & ... & Pn(tn,Mn,Cn,Dn) &
M1>L1 & ... & Mn>Ln & A1 .... & Am
->
P(t,M,f(M1,...,M2),g(C1,...,Cn,D1,...,Dn),union(D1,...,Dn))

Summary 2

Add evidence cumulating rule

Pred(X,M1,C1,D1) & Pred(X,M2,C2,D2) &
Empty(Intersection(D1,D2))
->
Pred(X,min(M1,M2),(C1+C2)-(C1*C2),union(D1,D2))

Add extended subsumption

Pred(X,M1,C1,D1) subsumes
Pred(Y,M2,C2,D2)
iff Y=Xs & M1>=M2 & C1>=C2 &
D1 is a subset of D2